CMO: Duality

Eklavya Sharma

1 Duality

Consider the optimization problem P:

$$\min_{x \in \mathbb{R}^d} f(x) \text{ where } \forall i \in I, c_i(x) \geq 0 \wedge \forall j \in J, h_j(x) = 0$$

The corresponding Lagrangian is

$$L(x, \lambda, \mu) = f(x) - \lambda^T c(x) - \mu^T h(x)$$

Define g as

$$g(\lambda, \mu) = \min_{x \in \mathbb{R}^d} L(x, \lambda, \mu)$$

Let D be this optimization problem:

$$\max_{\lambda, \mu} g(\lambda, \mu) \text{ where } g(\lambda, \mu) \neq -\infty \wedge \lambda \geq 0$$

Then D is said to be the dual of P.

Theorem 1 (Weak duality theorem). *Let x_0 be a feasible solution to P and (λ_0, μ_0) be feasible solution to D. Then*

$$g(\lambda_0, \mu_0) \leq L(x_0, \lambda_0, \mu_0) \leq f(x_0)$$

Proof.

$$\begin{aligned}
&g(\lambda_0, \mu_0) \\
&= \min_{x \in \mathbb{R}^d} L(x, \lambda_0, \mu_0) \\
&\leq L(x_0, \lambda_0, \mu_0) \\
&= f(x_0) - \lambda_0^T c(x_0) - \mu^T h(x_0) \\
&\leq f(x_0) && (\lambda_0 \geq 0 \wedge c(x_0) \geq 0 \wedge h(x_0) = 0 \text{ by feasibility})
\end{aligned}$$

□

Definition 1 (Duality gap). *Let x^* be the optimal solution to P and (λ^*, μ^*) be the optimal solution to D. Then the duality gap is defined to be the quantity*

$$f(x^*) - g(\lambda^*, \mu^*)$$

Corollary 1.1. *Let x_0 be a feasible solution to P and (λ_0, μ_0) be a feasible solution to D. If $f(x_0) = g(\lambda_0, \mu_0)$, then the duality gap is 0 and x_0 and (λ_0, μ_0) are optimal solutions.*

Proof. Let x^* be the optimal solution to P and (λ^*, μ^*) be the optimal solution to D. Then

$$g(\lambda_0, \mu_0) \leq g(\lambda^*, \mu^*) \leq f(x^*) \leq f(x_0) = g(\lambda_0, \mu_0)$$

Therefore,

$$g(\lambda_0, \mu_0) = g(\lambda^*, \mu^*) = f(x^*) = f(x_0)$$

□

2 Wolfe Dual

We'll now focus our attention on convex optimization problems. In the optimization problem P:

- Let f be a convex function.
- Let $c_i(x) = -f_i(x)$, where f_i is a convex function.
- Let $h_j(x) = a_j^T x - b_j$, where $a_j \in \mathbb{R}^d$ and $b \in \mathbb{R}^{|I|}$. Let A be the matrix whose j^{th} column is a_j.

Let WD be the optimization problem

$$\max_{x,\lambda,\mu} L(x, \lambda, \mu) \text{ where } \lambda \geq 0 \wedge \nabla_x L(x, \lambda, \mu) = 0$$

This problem is called the Wolfe Dual of P.

Theorem 2 (Proved previously)**.** *Let f be C^1 and convex. Then*

$$\forall u, v \in \mathbb{R}^d, f(v) \geq f(u) + \nabla_f(u)^T(v - u)$$

Lemma 3 (Proved previously)**.** *Let (x^*, λ^*, μ^*) be a KKT point. Then $f(x^*) = L(x^*, \mu^*, \lambda^*)$.*

Theorem 4. *Let (x^*, λ^*, μ^*) be a KKT point of P. Then (x^*, λ^*, μ^*) is the optimal solution to WD.*

Proof. Let (x, λ, μ) be a feasible point of WD.

$$\begin{aligned}
& L(x^*, \lambda^*, \mu^*) \\
& = f(x^*) && \text{(by lemma 3)} \\
& \geq L(x^*, \lambda, \mu) && \text{(by } \lambda \geq 0 \text{ and weak duality)} \\
& = f(x^*) + \sum_i \lambda_i f_i(x^*) + \sum_j \mu_j (a_j^T x^* - b_j) \\
& \geq (f(x) + \nabla_f(x)^T (x^* - x)) \\
& \quad + \sum_i \lambda_i (f_i(x) + \nabla_{f_i}(x)^T (x^* - x)) \\
& \quad + \sum_j \mu_j (a_j^T (x^* - x) - (a_j^T x - b_j)) && \text{(by theorem 2)} \\
& = \left(f(x) + \sum_i \lambda_i f_i(x) + \sum_j \mu_j (a_j^T x - b_j) \right) \\
& \quad + (x^* - x)^T \left(\nabla_f(x) + \sum_i \lambda_i \nabla_{f_i}(x) + \sum_j \mu_j a_j \right) \\
& = L(x, \lambda, \mu) + (x^* - x)^T (\nabla_x L(x, \lambda, \mu)) \\
& = L(x, \lambda, \mu) && \text{(feasibility of WD implies } \nabla_x L(x, \lambda, \mu) = 0)
\end{aligned}$$

Therefore, (x^*, λ^*, μ^*) maximizes WD. □

Therefore, to find the KKT point of a problem, we can optimize its Wolfe Dual.

Example 1.

$$\min_x \frac{1}{2} \|x\|^2 \ \textit{ where } A^T x \geq b$$

The Lagrangian for this problem is

$$L(x, \lambda) = \frac{1}{2} \|x\|^2 - \lambda^T (A^T x - b)$$

$$\nabla_x L(x, \lambda) = x - A\lambda$$

The Wolfe Dual is

$$\max_{x,\lambda} \frac{1}{2} \|x\|^2 - \lambda^T (A^T x - b) \ \textit{ where } x - A\lambda = 0 \textit{ and } \lambda \geq 0$$

We can simplify this by substituting $x = A\lambda$ and removing the constraint

$$\max_\lambda b^T \lambda - \frac{1}{2} \|A\lambda\|^2 \ \textit{ where } \lambda \geq 0$$

Example 2.

$$\min_x c^T x \ \textit{ where } x \geq 0 \wedge Ax \geq b$$

The Lagrangian for this problem is

$$L(x, \lambda, \pi) = c^T x - \lambda^T (Ax - b) - \pi^T x = (c - A^T \lambda - \pi)^T x + b^T \lambda$$

$$\nabla_x L(x, \lambda, \pi) = c - A^T \lambda - \pi$$

The Wolfe Dual is

$$\max_{x,\lambda,\pi} (c - A^T \lambda - \pi)^T x + b^T \lambda \text{ where } c - A^T \lambda - \pi = 0 \text{ and } \lambda \geq 0 \text{ and } \pi \geq 0$$

We can simplify this by substituting $\pi = c - A^T \lambda$ *and removing the constraint*

$$\max_{x,\lambda} b^T \lambda \text{ where } A^T \lambda \leq c \text{ and } \lambda \geq 0$$

This gives us the dual linear program for this problem.